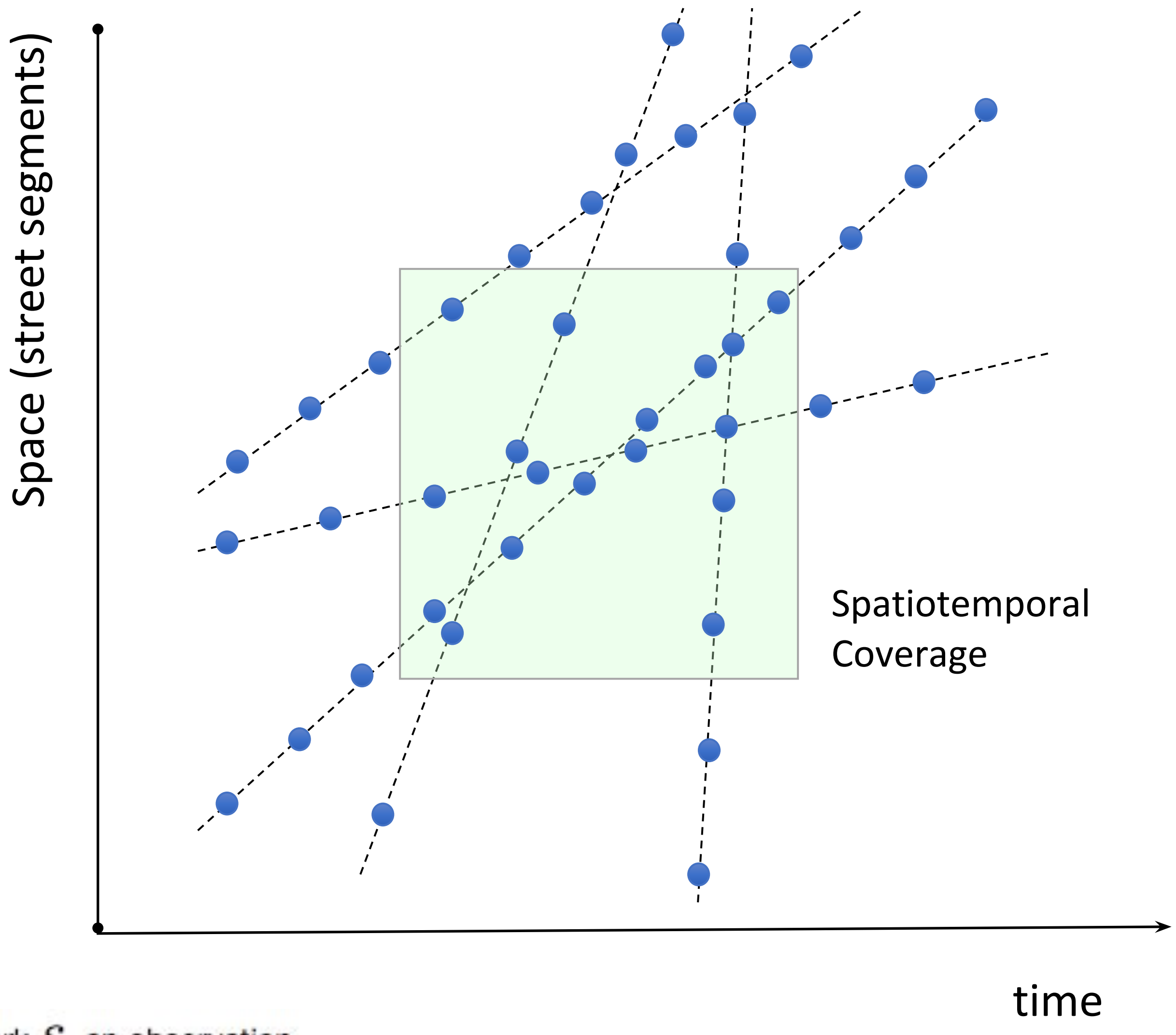
City Scanner-

Low-cost mobile air quality monitoring using trash trucks in the city of Cambridge, Massachusetts

- Priyanka deSouza
Senseable City Lab, MIT

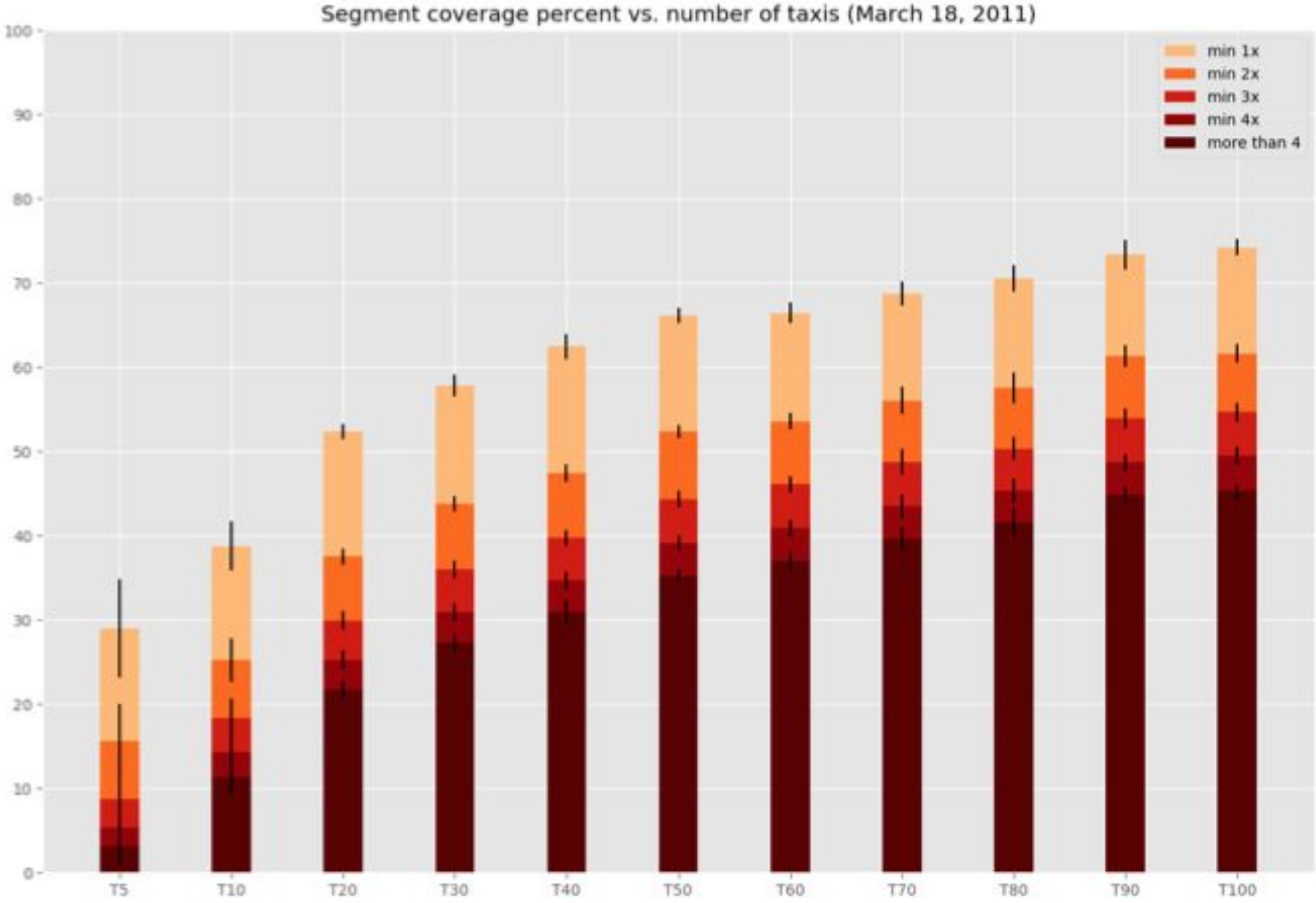

FREIGHTLINER
CAMBRIDGE
DEPARTMENT
PUBLIC

Mobile Sensors



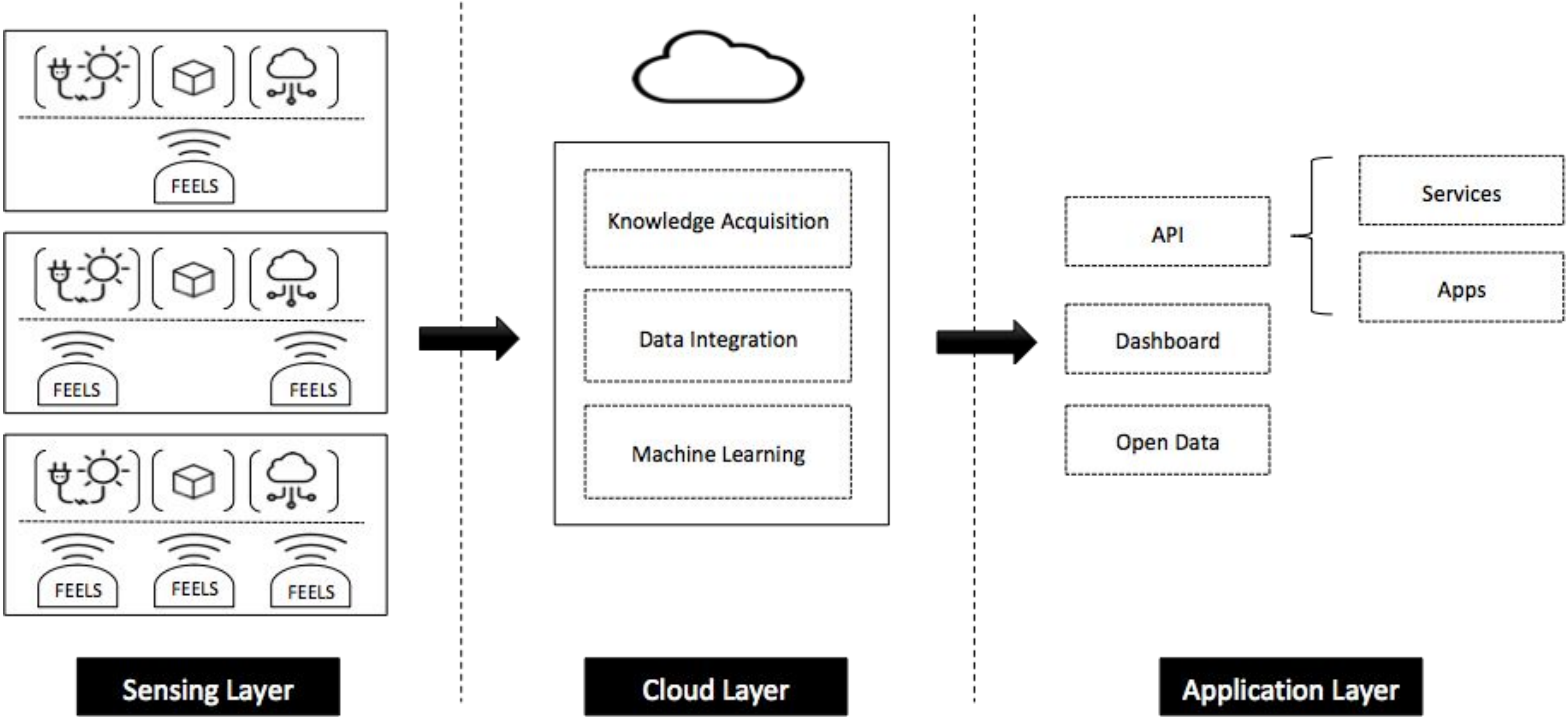
MINIMUM STREET SAMPLING PROBLEM: Given a street network S, an observation period T, a minimum sampling requirement $\overline{n}$, and a collection $\mathcal{V}$ of vehicles moving in S during T according to some mobility pattern $\mathcal{M}$, what is the minimum number of vehicles randomly selected from $\mathcal{V}$ that ensures that each element in S is covered by at least $\overline{n}$ vehicles?

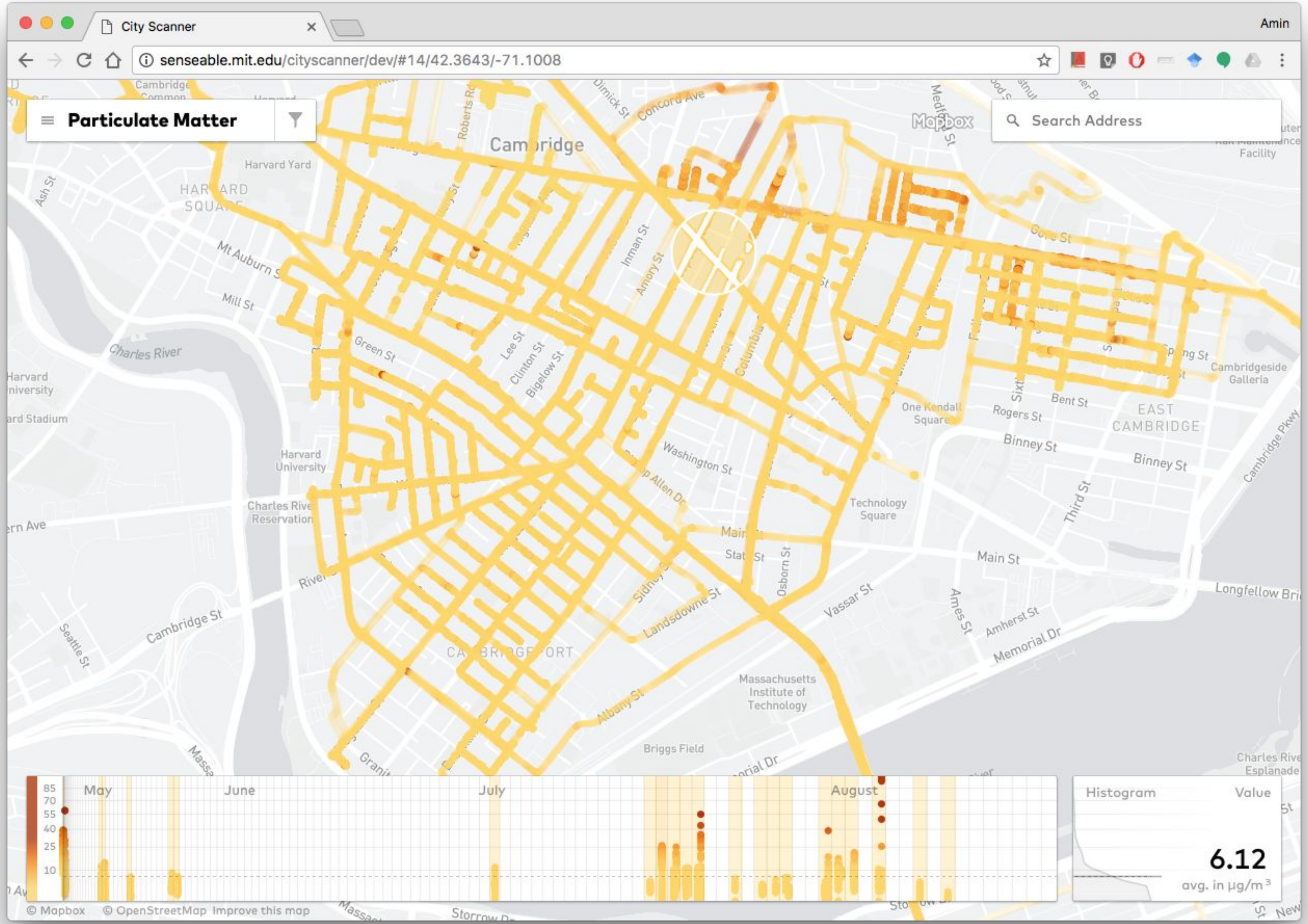
How many vehicles do we need to guarantee a given sensing quality?



Anjomshoaa, A., Duarte, F., Rennings, D., Matarazzo, T., de Souza, P., & Ratti, C. (2018). City Scanner: Building and Scheduling a Mobile Sensing Platform for Smart City Services. IEEE Internet of Things Journal.

System Architecture



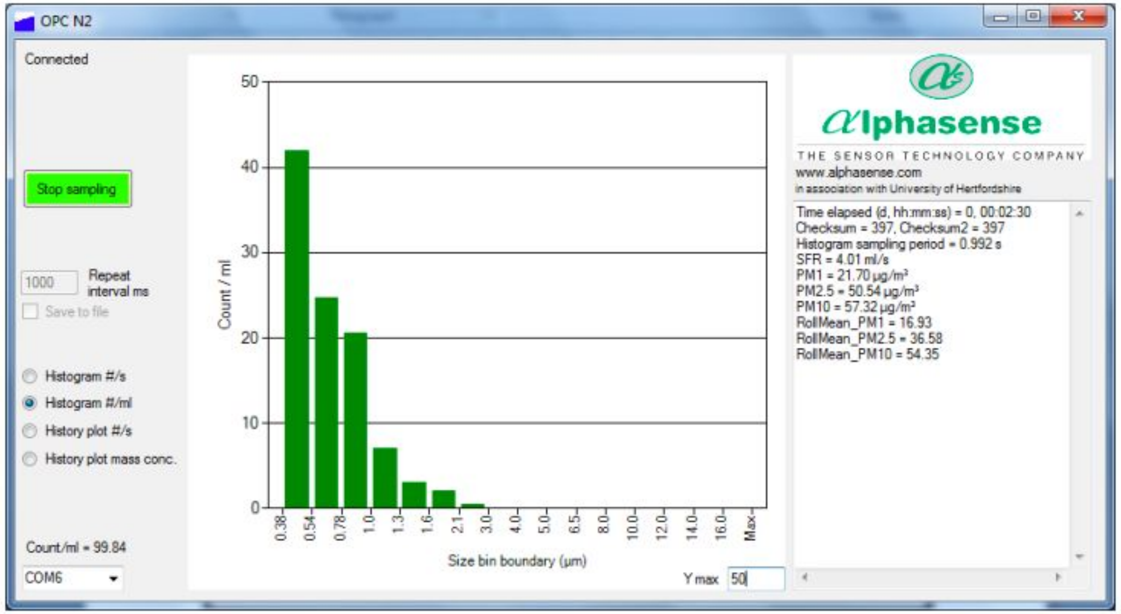


http://senseable.mit.edu/cityscanner/

Making sense of the air pollution data

Sensor type: Alphasense OPC-N2. Frequency of sampling: 60 Hz

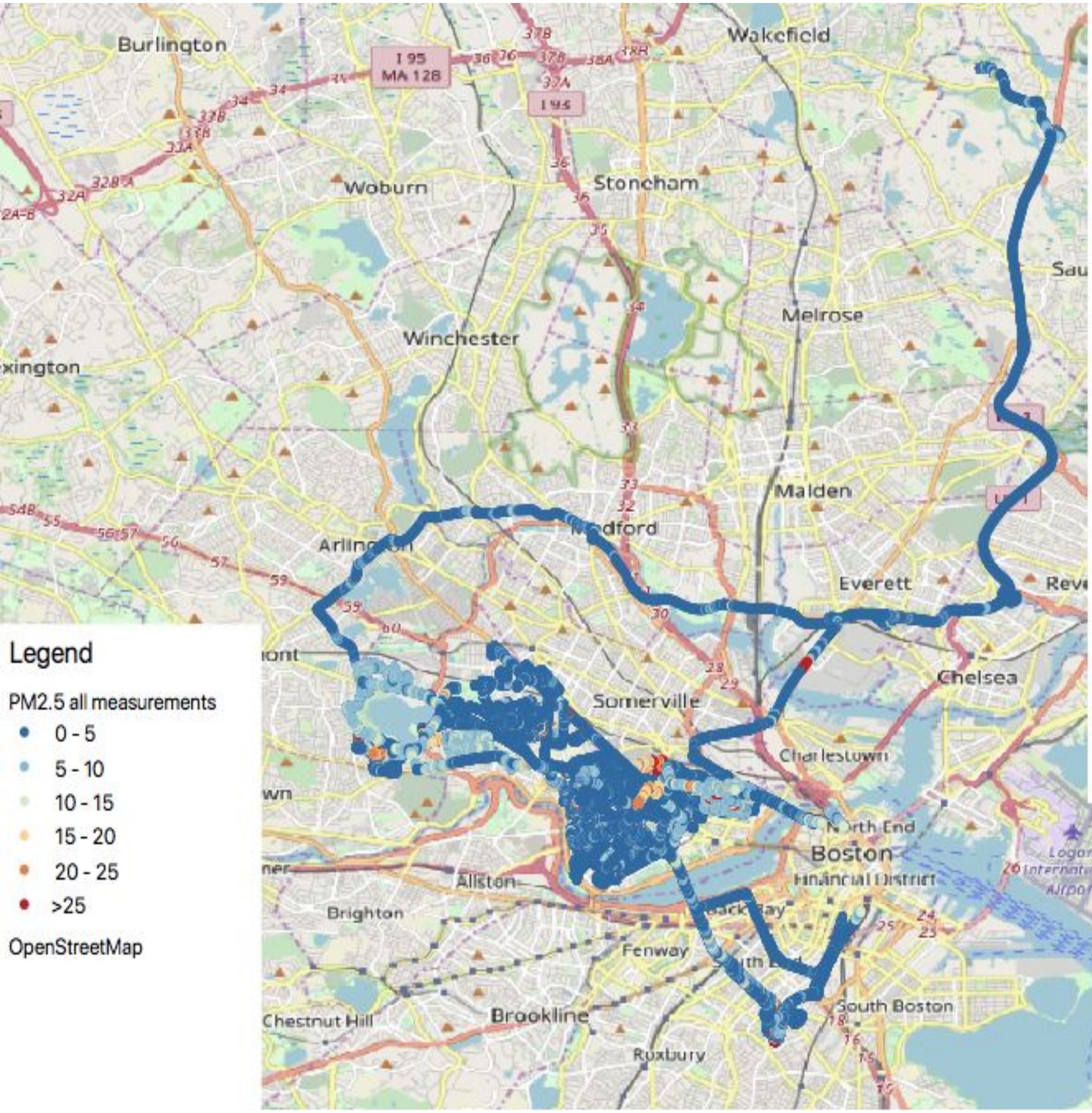
It measures the number concentration of particles in 16 size bins between 0.38 and 17.5μm.



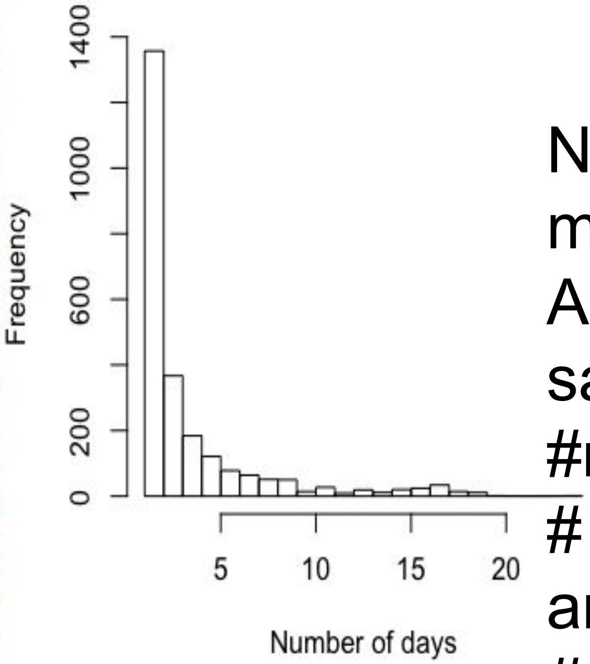
Sousan et al., (2016) say that the OPC under-reports the number of particles having diameters <0.8 μm (detection efficiency: 78%). The detection efficiency for the other sizes is between 80-110%

Making sense of the air pollution data

The trash-trucks did not cover the same route for each run.

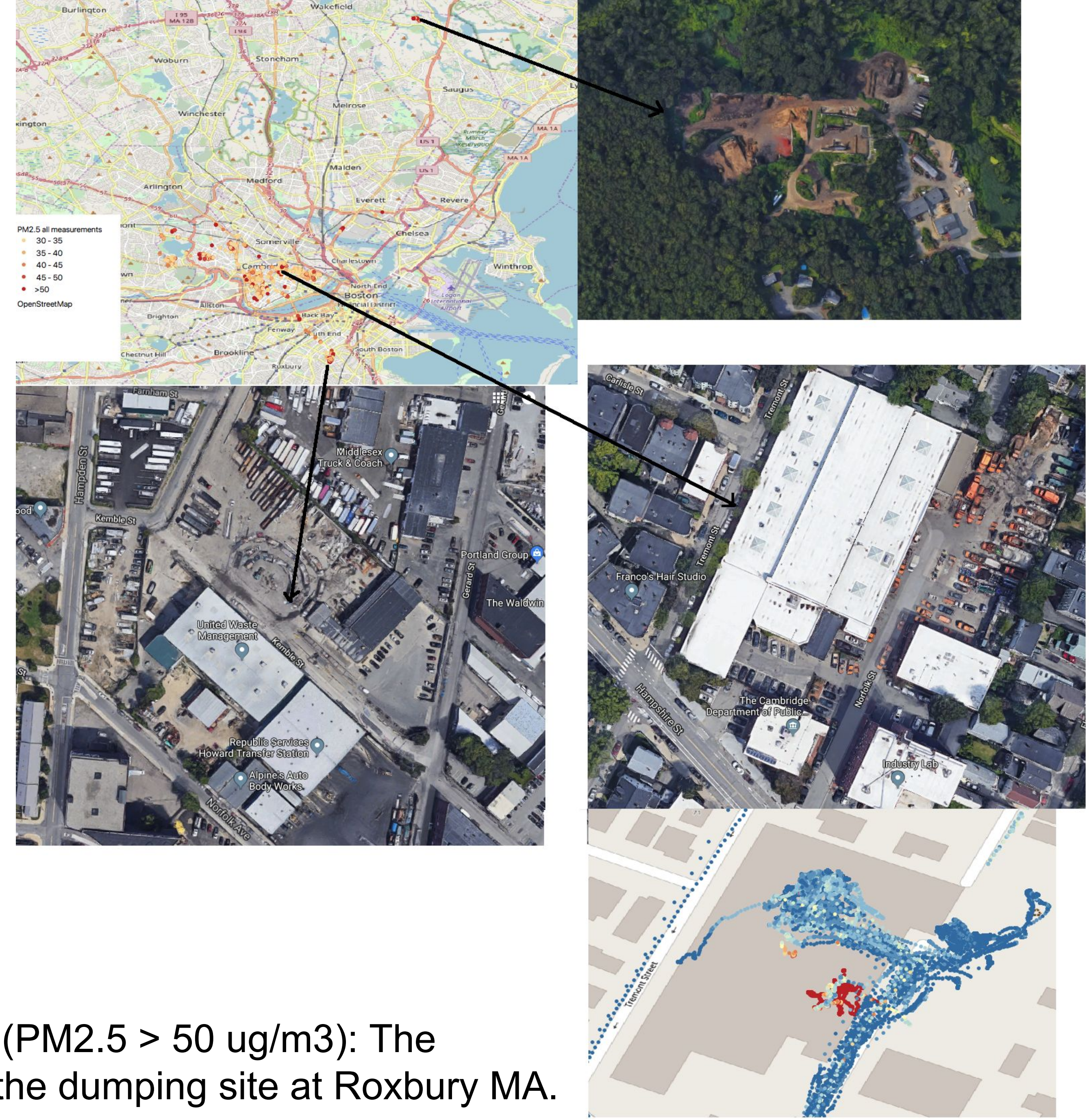


Geographic extent of all the runs



Number of unique days of measurement: 27
Avg frequency for a road to be sample: 4
#roads sampled>15 times: 105
roads sampled between 10 and 15: 80
roads sampled<10 times: 2272

Hot Spots?



From the raw data we noticed three persistent hotspots (PM2.5 > 50 ug/m3): The Trashtruck depot, the dumping site at Saugus, MA and the dumping site at Roxbury MA. Features of the kind of vehicle we have chosen?

Comparing air pollution across Cambridge

1) We divide the total routes that the trash trucks cover into segments of 30 meters (small enough to have enough air pollution points/ segment, and large enough such that the error in the GPS device (6 meters) does not matter

2) In order to compare air pollution for each segment across space, we had to normalise over time. We do this by estimating the diurnal pattern of air pollution and normalising our measurements, using this trend

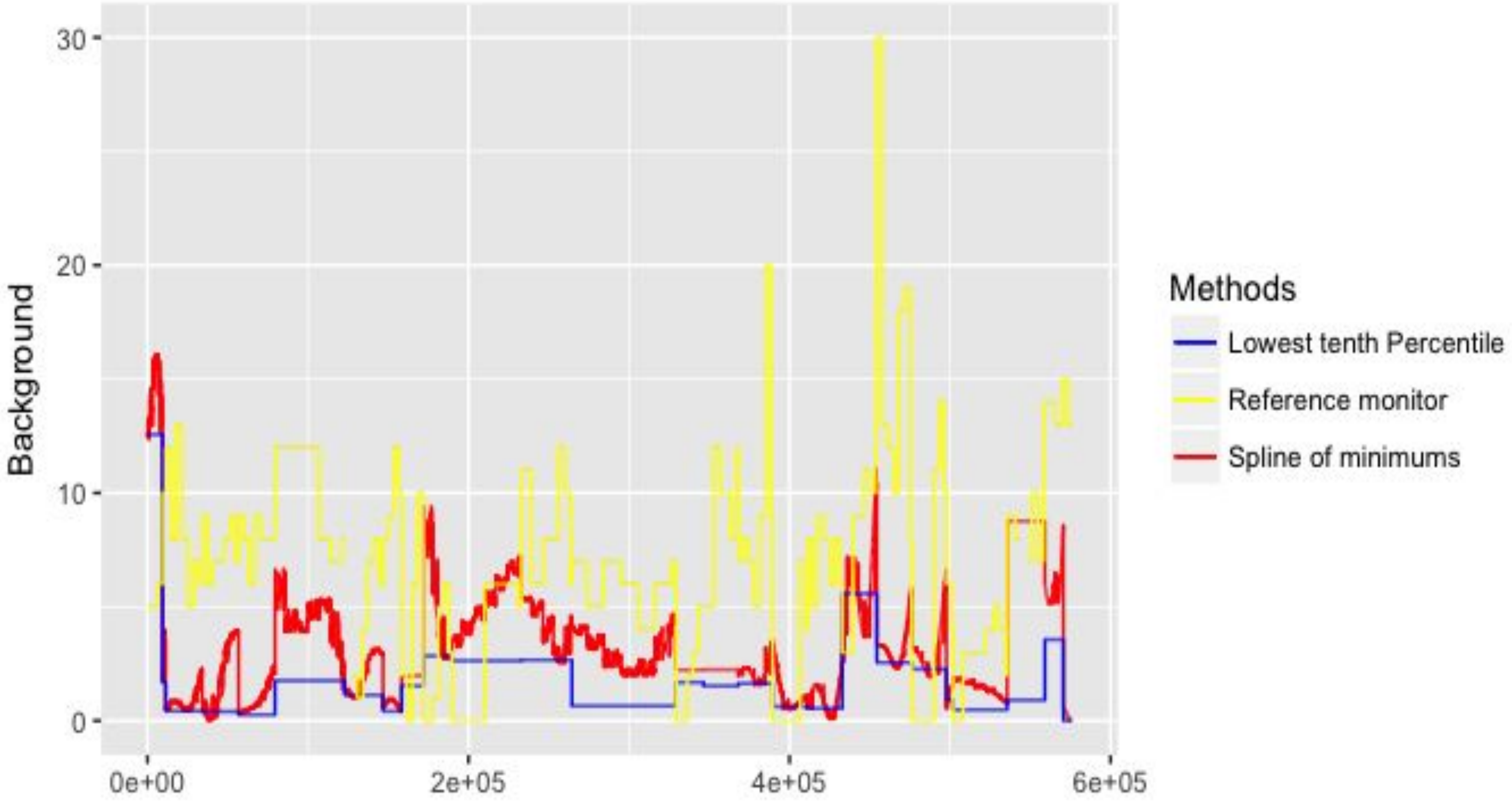
$PM_{2.5,norm\ i}$= $PM_{2.5,OPC\ i}$ x (median $PM_{2.5,\ background}$ value for the time period 7 am to 2 pm on the day of measurement/$PM_{2.5,\ background\ i}$).. (2)

where $PM_{2.5,OPC\ i}$ is the OPC measurement for event i,

However, as Cambridge can be considered a city with clean air, often the PM2.5 background calculated is very low. In this case, we made use of an additive multiplicative factor instead of a multiplicative factor using

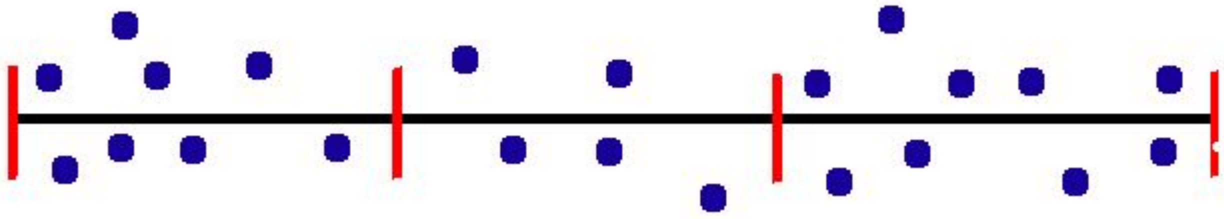
$PM_{2.5,norm\ i}$= $PM_{2.5,OPC\ i}$ - $PM_{2.5,\ background\ i}$ + (median $PM_{2.5,\ background}$ value for the time period 7 am to 2 pm on the day of measurement).. (3)

Background correction



In order to compare measurements made at the same location, but on different days and at different times, we need to account for possible bias created by diurnal variation in urban-background pollution

Comparing air pollution across Cambridge



3) We then snap our normalised measurements to the 30 meter road segment that the measurement was made on

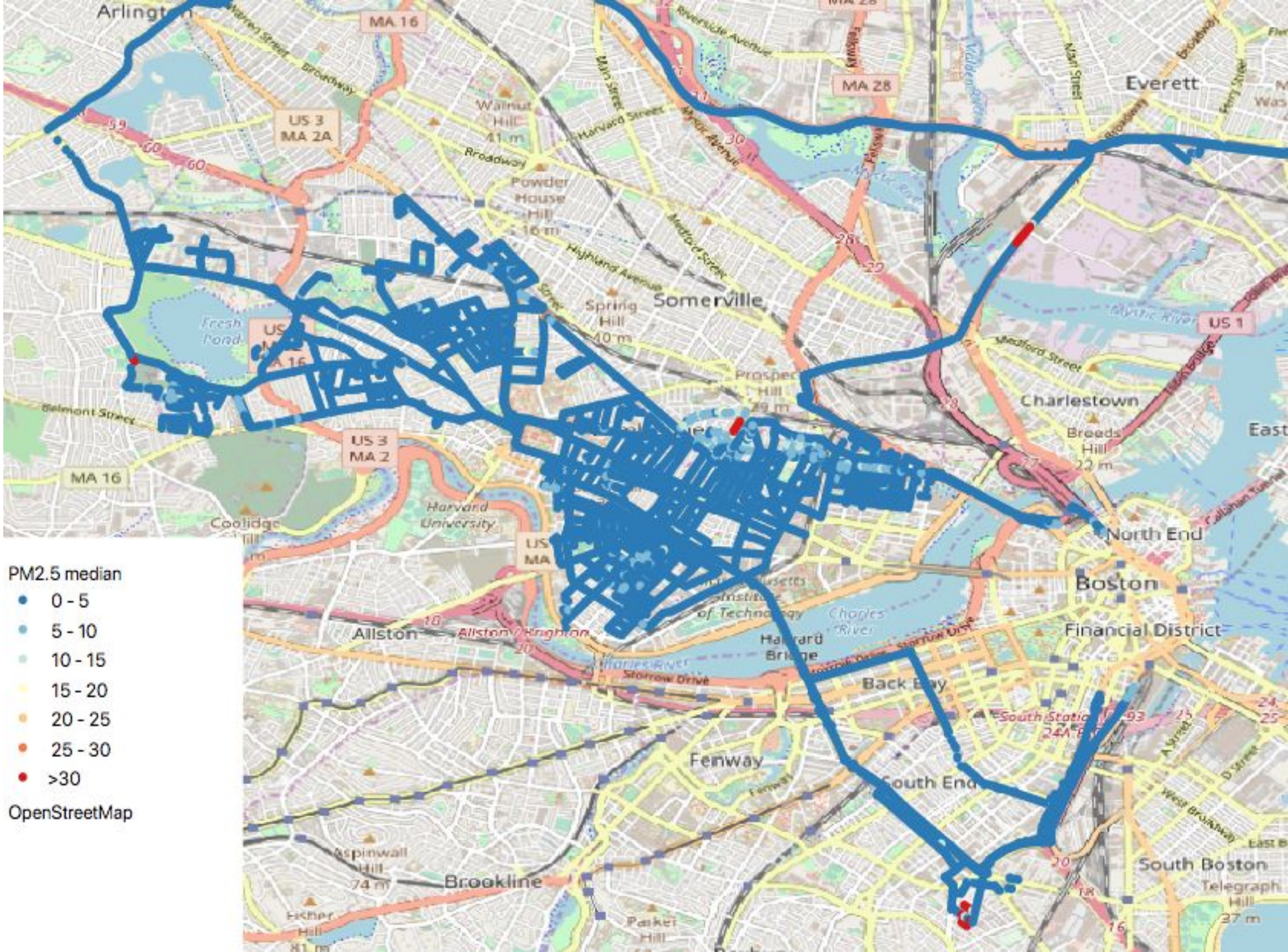
4) We use the median PM2.5 value of the segment as the ‘generalizable’ PM2.5 value of that 30 meter segment. We choose the median to avoid considering outliers

5) We then estimate the bias/RMSE and spatial and temporal stability of our results by evaluating the intra-cluster correlation coefficient

Background corrected median PM2.5 for each road segment

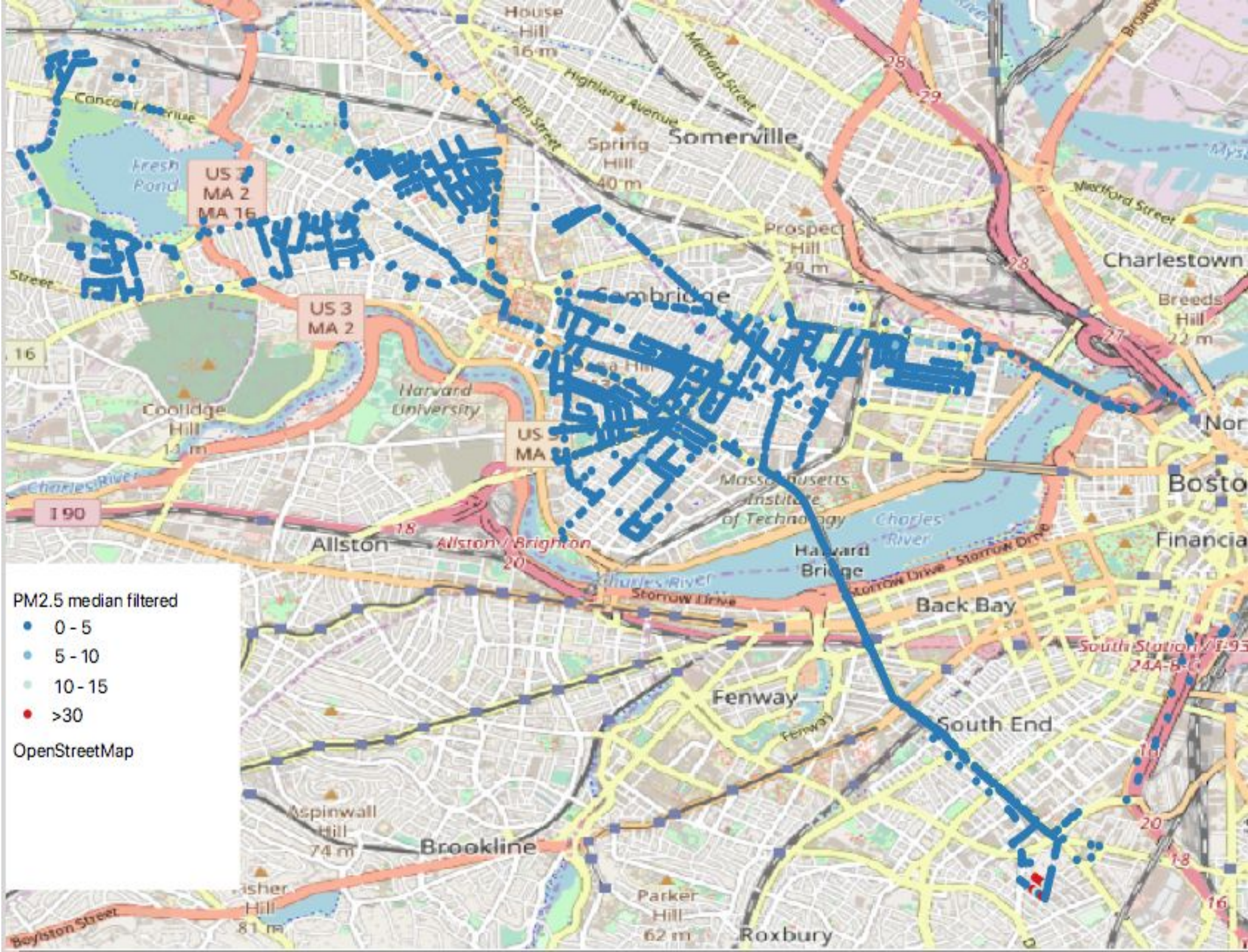
PM2.5 is low in Cambridge for the time period of measurement by the trash truck 7 am-2 pm on weekdays.

Function of trucks being on the street when lots of traffic is not out?

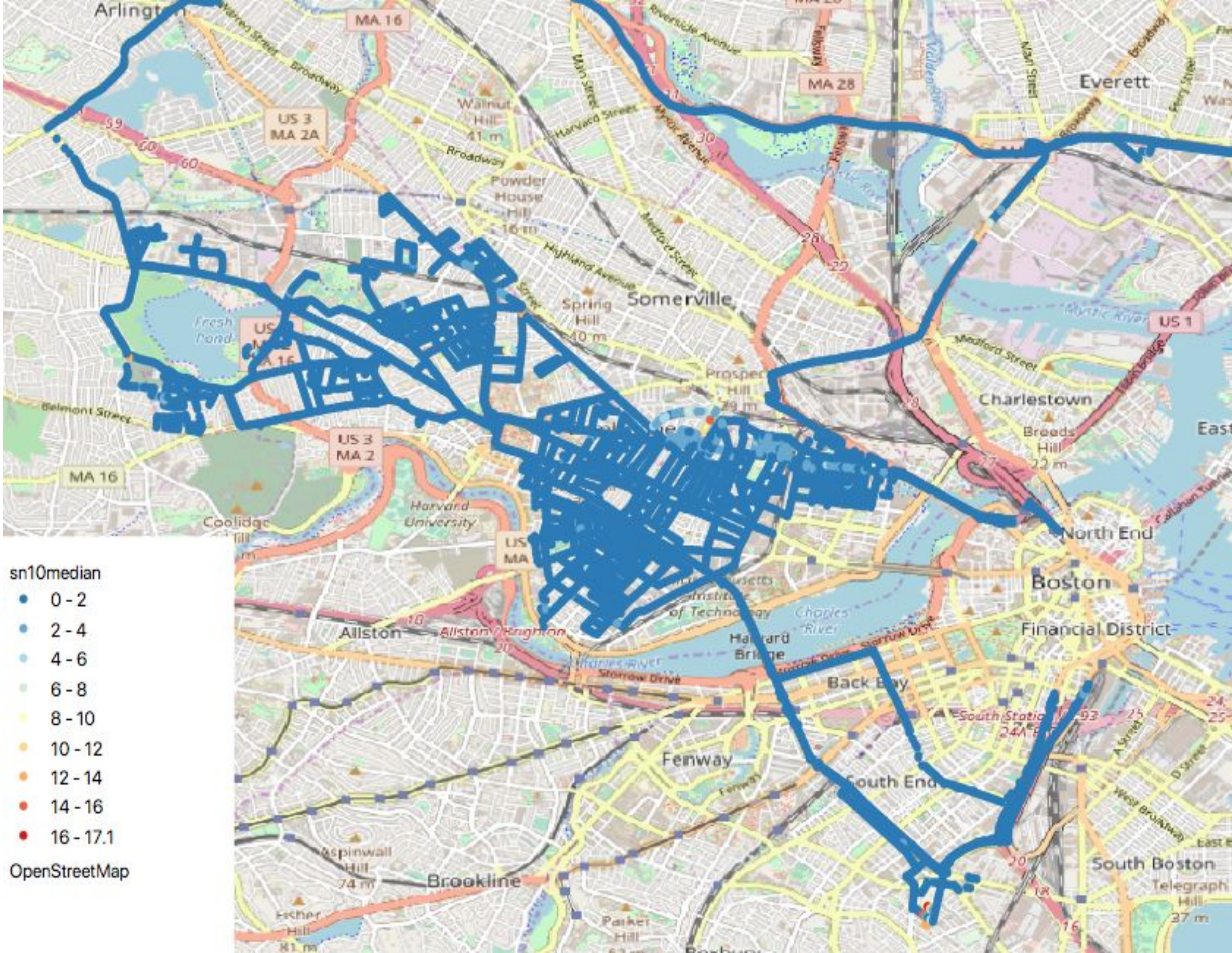


Background corrected median PM2.5 for each road segment

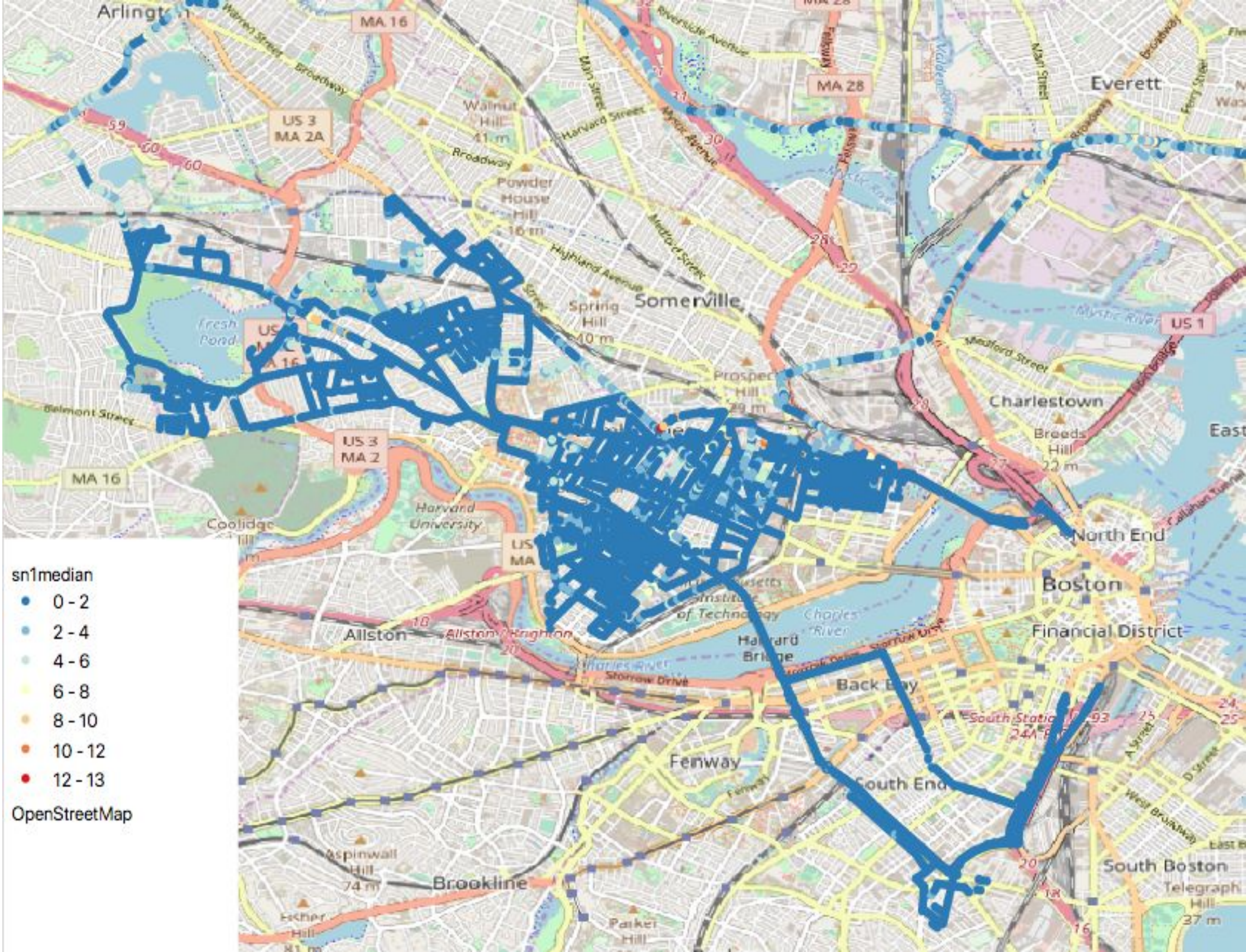
For segments where the normalized error < 0.2 and days sampled > 1

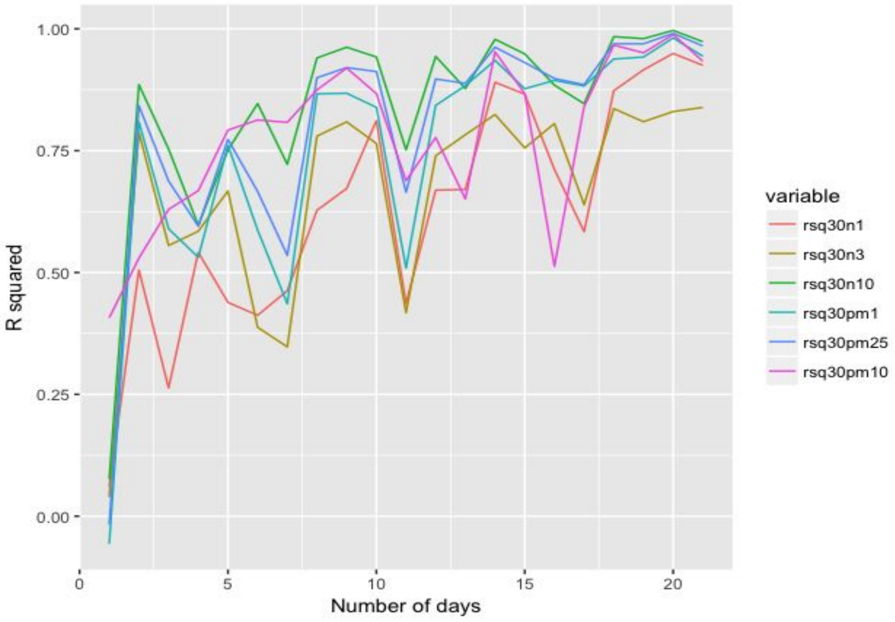


Background corrected number concentration of particles having diameters > 1 μm < 10 μm

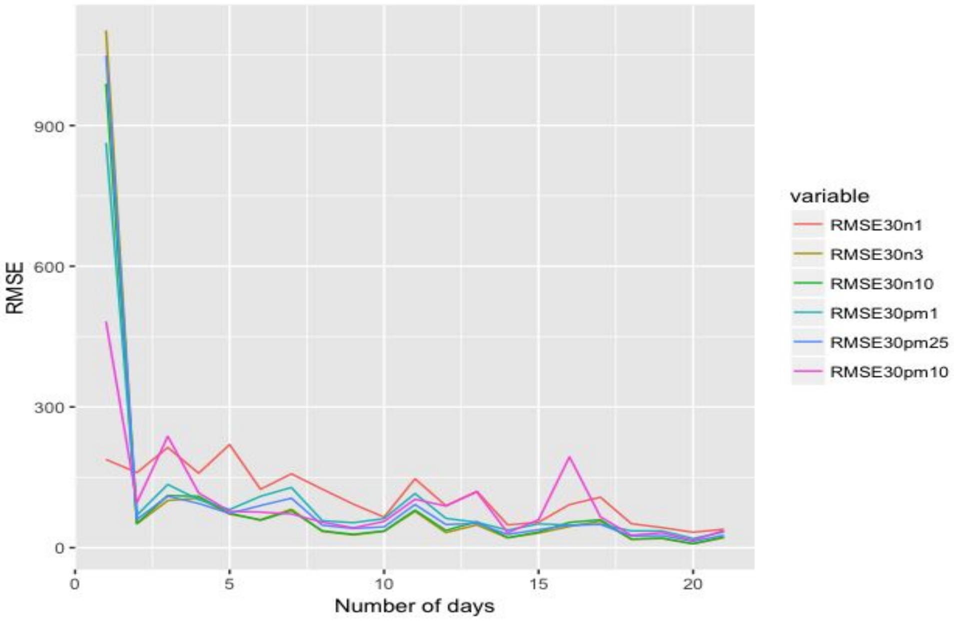


Background corrected number concentration of particles having diameters > 0.38 µm < 1 µm

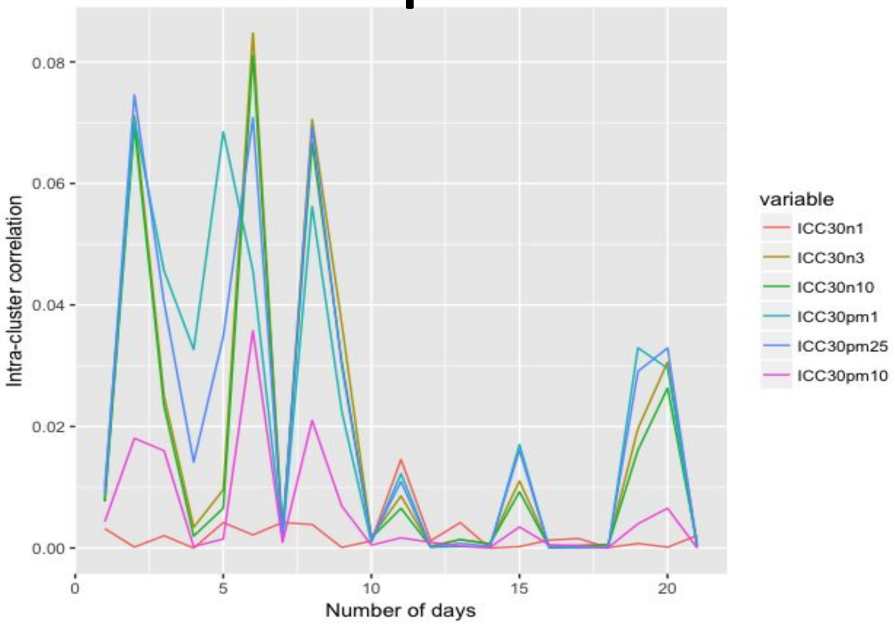




Metric of precision



Metric of bias



Metric of 'uniqueness'

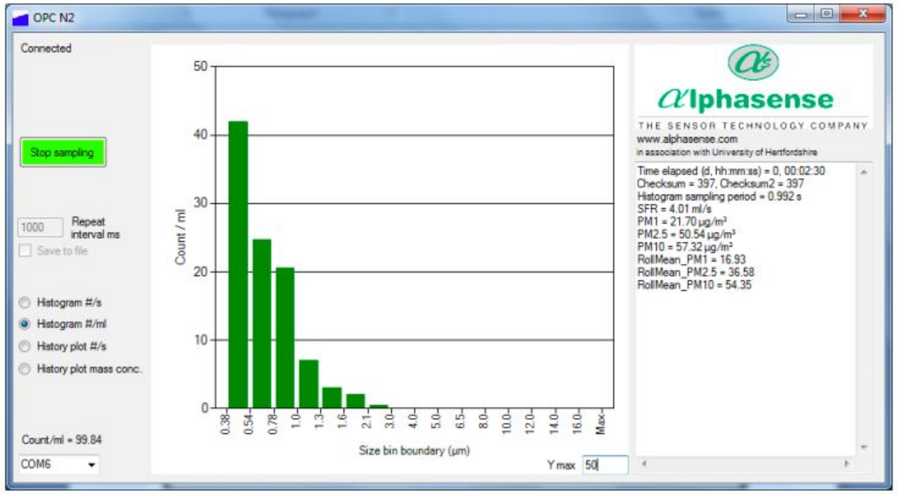
We used a Monte Carlo subsampling technique where we sampled observations for a subset of days ranging from 1:24 days

Using particle count information from the OPCs

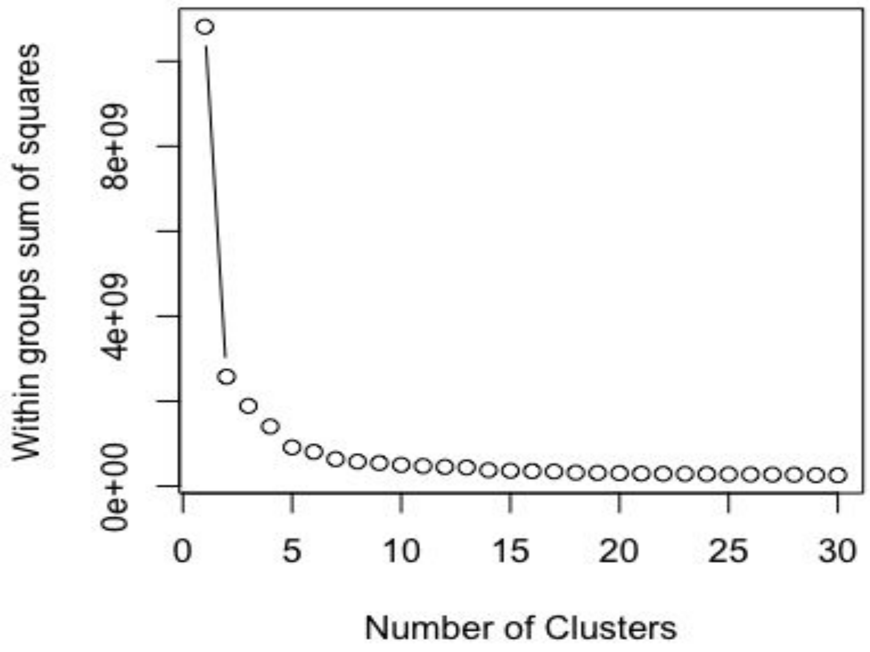
In addition to PM2.5, the OPC also outputs raw particle counts in 16 diameter bins. This means we have 16 data points/measurement

In order to analyse this data, we use the k-means clustering technique to cluster the the raw particle count data into groups

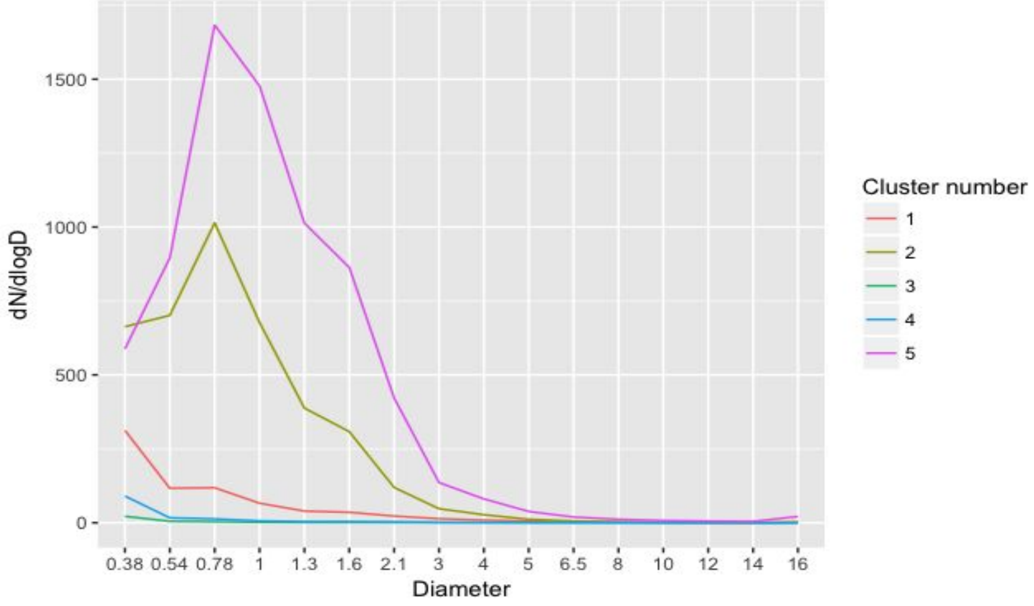
We will then analyse the average properties of each group in order to gain insights into the particle formation processes.



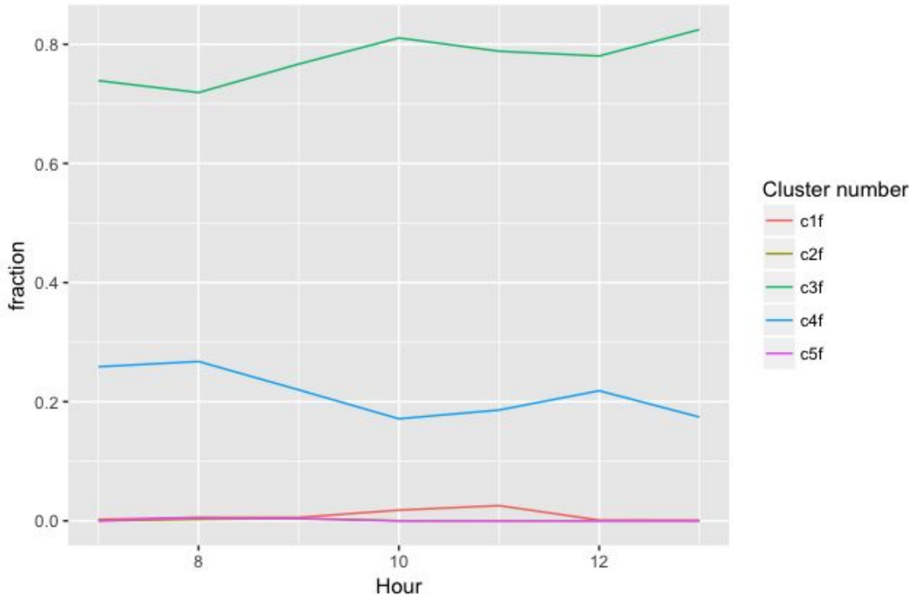
Making sense of the 16 number concentrations for different particle size bins for each observations



We see that the within-group variance decreases dramatically after 5 clusters. We thus for the sake of interpretation choose 5 clusters



Average size distribution for each cluster



Average hourly fraction of each cluster

Cluster properties

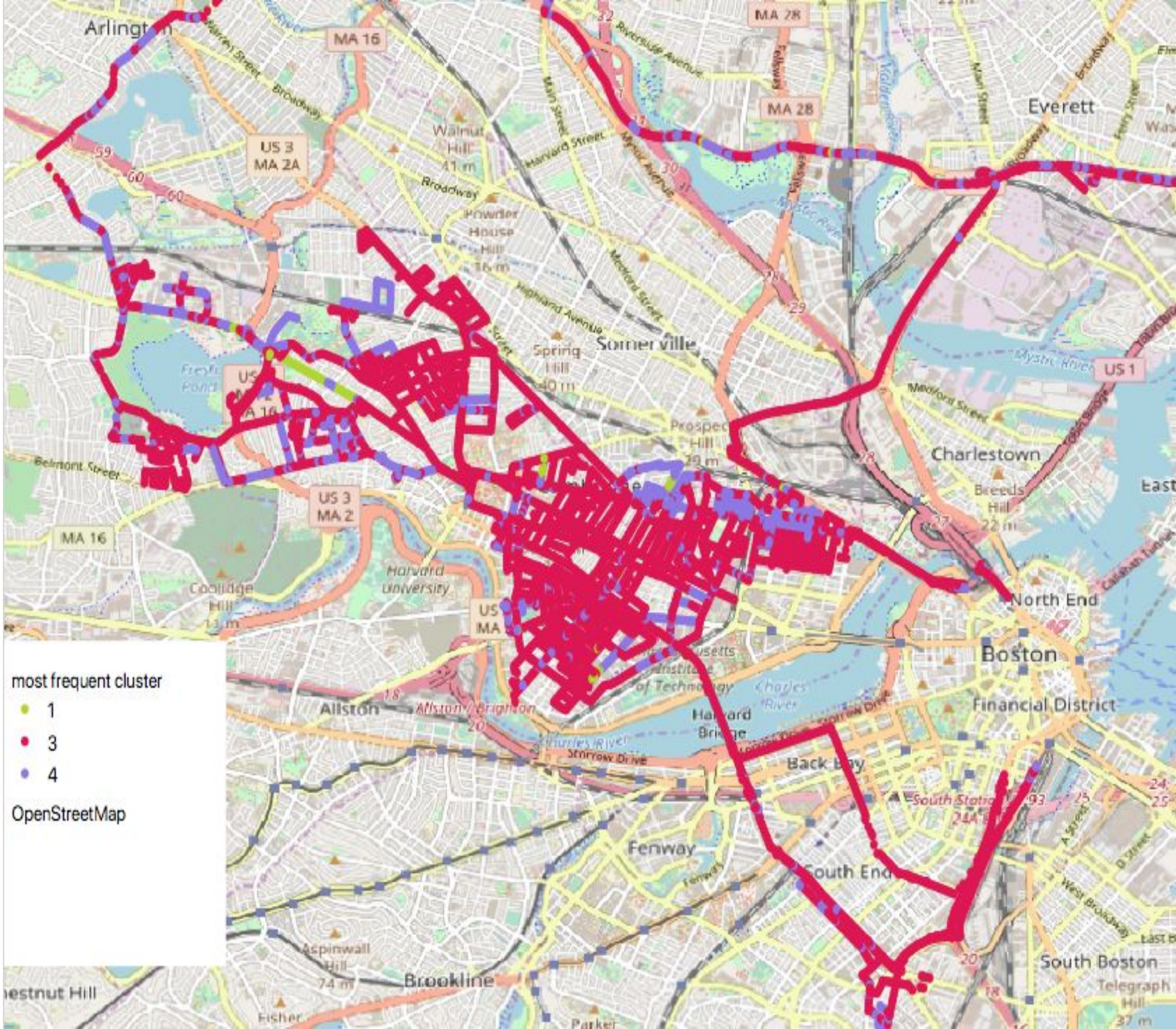
Cluster number	PM1 µg/m^3	PM2.5 µg/m^3	PM10 µg/m^3	N1 #/cm^3	N10 #/cm^3	Number of days	Average velocity (m/s)	Number of observations corresponding to each cluster	Background contribution to PM2.5 measured
1	41	70	315	79	22	24	1.5	5039 (0.87%)	8.2%
2	215	419	817	322	178	7	0.07	682 (0.12%)	2.19%
3	2	3	13	4	1	27	1.17	440,988 (76.6%)	3.13%
4	8	11	38	18	2	27	1.43	128,205 (22.3%)	50.4%
5	321	804	2586	414	460	2	0.13	886 (0.5%)	1.1%

Correlation between PM2.5 and velocity= - 0.023

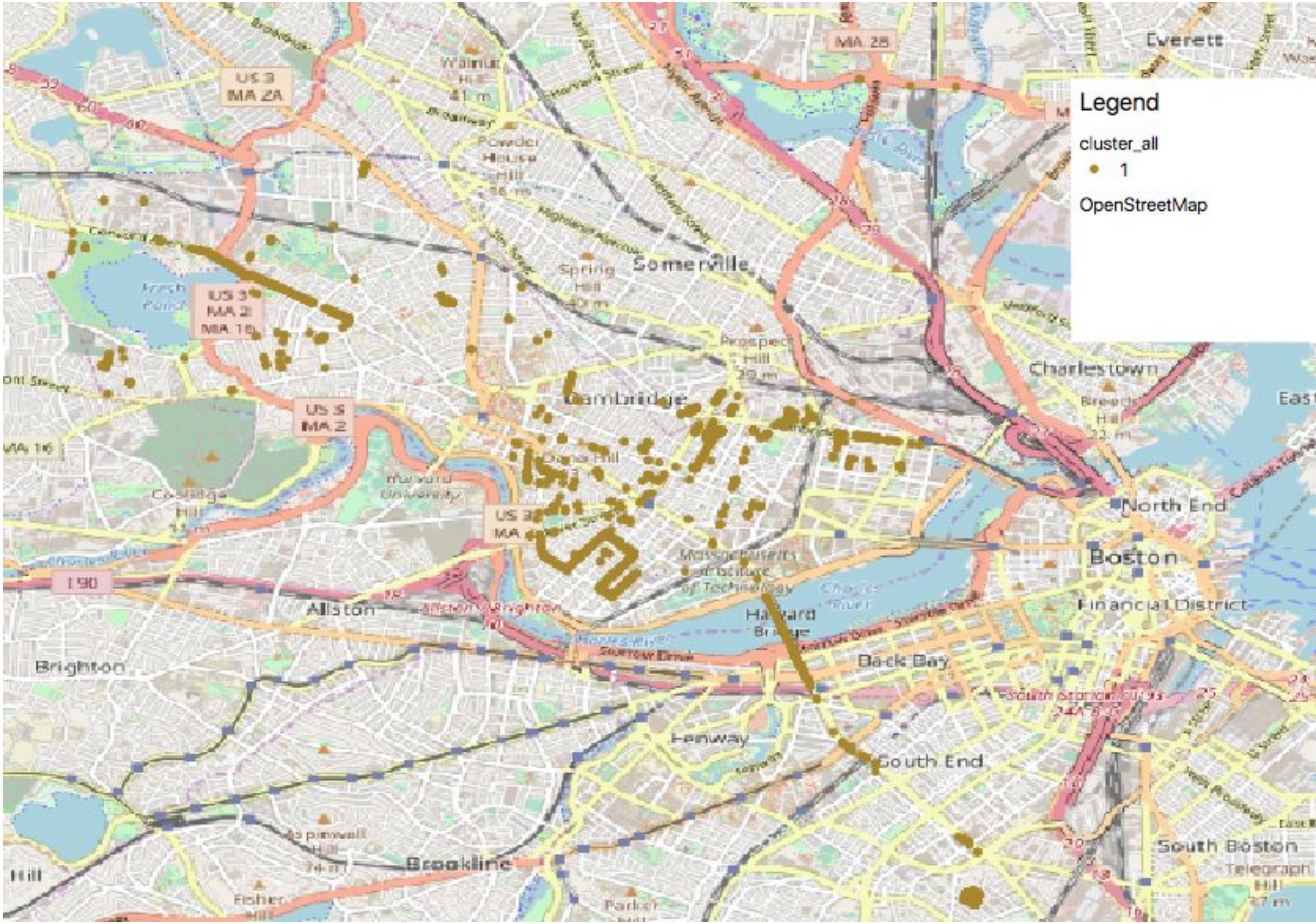
Cluster 4 has a higher percentage of background air pollution. The difference between cluster 4 and 3 may only be an artefact of the different times in the day that measurements were made.

However, the high number concentrations corresponding to 2 and 5 especially indicate they correspond to local sources. Identifying these sources is an important next step

Most frequent cluster number on each road segment

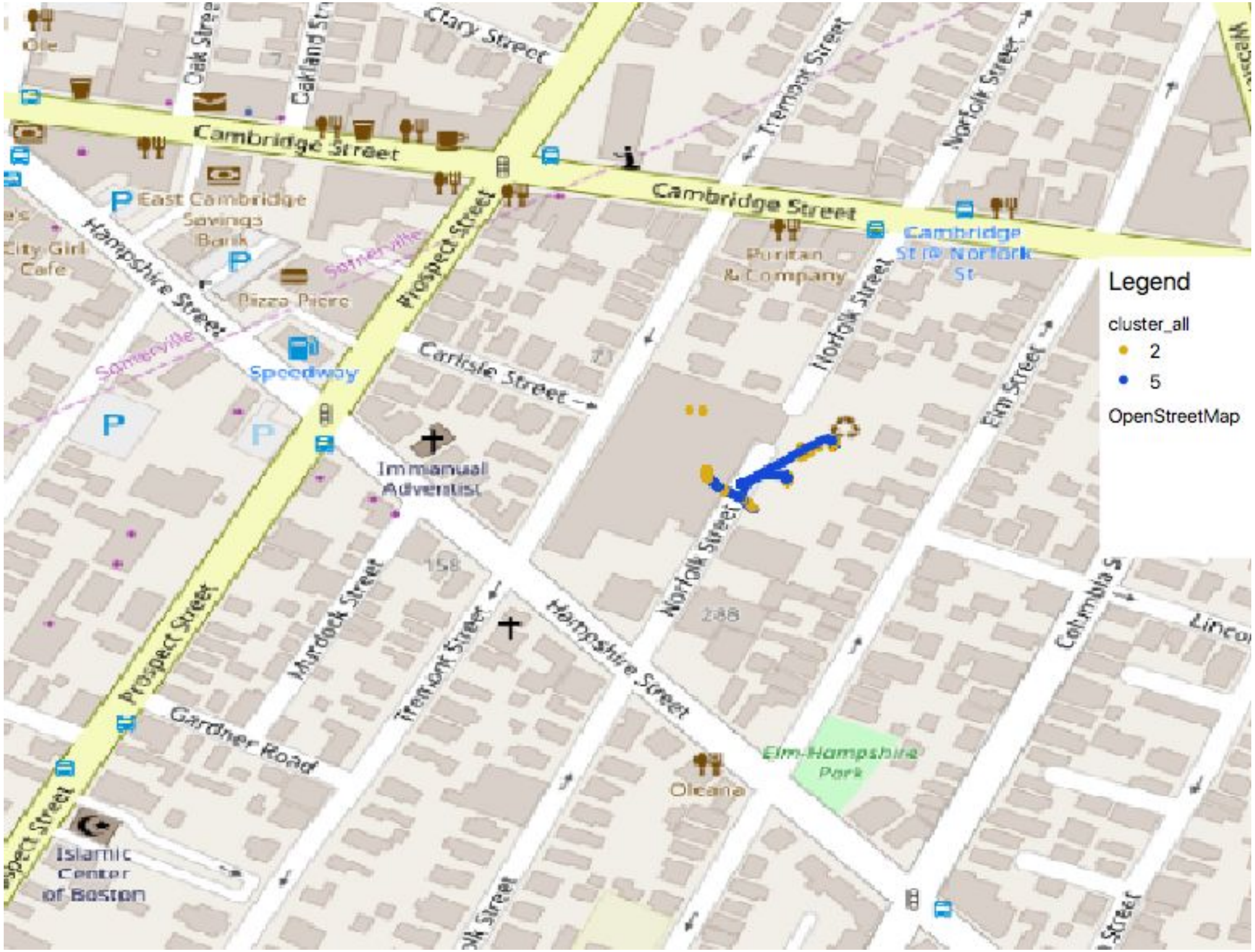


Cluster 1



Legend
cluster_all
1
OpenStreetMap
Somerville
Cambridge
Charlestown
North End
Boston
Financial District
Back Bay
South End
Fenway
Allston
Brighton
Brookline
South Boston
Everett
Harvard Bridge
Massachusetts Institute of Technology
Harvard University
Fresh Pond
Spring Hill
Prospect Hill
Powder House Hill
Walnut Hill
Dana Hill
Aspinwall Hill
Fisher Hill
Breeds Hill
Telegraph Hill
Parker Hill
US 3
MA 2A
MA 2
MA 16
I 90
MA 28

Cluster 2 and 5



Summary

1) PM2.5 concentrations in Cambridge are low on average between times of 7 am and 2 pm

2) Our measurements can be complemented by those on fixed monitors/on vehicles with different spatio-temporal characteristics

3) Our cluster analysis gives us some idea of air pollution sources in Cambridge. Our future deployment needs to delve into these sources in more detail